

## 5 Hybrid Libraries

**Karen Calhoun**

Cornell University Library (retired)

[ksc10@cornell.edu](mailto:ksc10@cornell.edu)

**Note:** This is a preprint of a chapter whose final and definitive form was co-published in *Exploring Digital Libraries: Foundations, Practice, Prospects* by [Facet Publishing](#) (2014) and [ALA Neal-Schuman](#) (2014).

### Overview

This chapter continues the discussion of digital collections with a detailed look at the interplay between library users, hybrid library collections and enabling technologies for hybrid library systems and services. Hybrid library collections contain non-digital, digitized and born digital resources. This chapter examines changing information-seeking behaviors and preferences, explores how they have fostered new collections strategies, and analyses the impact of both on discovery services and other enabling technologies for hybrid libraries. The chapter ends with some thoughts about the parallel but separate evolutionary paths of hybrid libraries, repositories and the web.

### Changing information-seeking behaviors

#### *Information moves online*

The content of interest to those who use libraries is highly distributed across the web. Vast changes have occurred not only in the amount of information available but also where people prefer to look for what they need. Library collections exist alongside (and compete for attention with) many other choices for information seekers, including those for whom hybrid library collections are or would be useful.

**Keywords:** Hybrid libraries; Information seeking; Library use; Collection management (Libraries); Mass digitization; Cultural heritage collections; E-resource management; Shared print repositories; Library cooperation; Integrated library systems; Discovery systems and services; Discoverability

Digital formats are beginning to dominate library collections, especially in academic libraries. Particularly with respect to the scholarly journal literature, library collections are already *digital* collections, and online formats are preferred. As discussed in chapter 2, by 2001 a third of faculty and half of students reported they were relying exclusively or almost exclusively on online scholarly resources for their work (Friedlander 2002). More than a decade later, preferences for web-based scholarly content are much stronger.

### ***Research on information-seeking behaviors***

#### *Preferred sources of information*

The attention of both the general public and academics has shifted rapidly to online networked content. Many people now prefer to look for information online, and most segments of the population place a high value on immediately available, convenient online sources, often preferring these sources over hybrid library collections. Much research has been focused on these trends, for example the following studies:

- **The American public.** According to a survey of people's perceptions of libraries and preferences for information discovery conducted by Harris Interactive on behalf of OCLC, 84% of surveyed Americans say they prefer to begin a search for information with a search engine. Furthermore, a majority (69%) of American respondents considered the information they find on the web to be as trustworthy as information from a library (De Rosa et al. 2011, 32, 40).
- **The British public.** Bob Usherwood reported on the results of a national survey to assess the value that the British public places on libraries, archives and museums as repositories of knowledge (2005). His findings suggest that libraries are still valued for their role as trusted sources of information, but the findings also confirm the trend found in other studies: a

preference for immediately accessible, convenient sources of information (the web, newspapers, television). Survey respondents also saw libraries' growing use of digitization and e-resources as positive steps for increasing what libraries can offer to an online world.

- **Undergraduates.** Head and Eisenberg (2010, 7) reported the results of their studies of the information-seeking behaviors of US undergraduates and the sources they consult for their coursework. Their study indicated that in 2010 the top three sources used by undergraduates for completing coursework were course readings (96%), search engines (92%), and online scholarly resources (88%). Students also frequently used Wikipedia to support their coursework (73%).
- **US and UK faculty.** An Ithaka longitudinal study of US faculty members' preferences for starting their research suggests that most begin with a discipline-specific e-resource (over 40%) or with a search engine (about 35%). Less than 20% begin with library online catalogs. These trends held up across respondents from the social sciences and sciences disciplines, with humanists showing roughly equal preference for starting research with discipline-specific e-resources, search engines and the online catalog (Schonfeld, Housewright and Wulfson 2013, 21-22). The study was repeated in the UK; results indicated that 40% of UK faculty members begin their research with a search engine, 33% with a discipline-specific e-resource, and 15% each with an online or national/international library catalog (Housewright, Schonfeld and Wulfson 2013, 21-22).

#### *Web referral traffic and destinations*

Web referral traffic comes from external web sites and pages (these are called "referrers") that lead web users to another site or page (these are called "destinations," in this context, digital library sites with specific URLs). In July 2010, one web technology analyst (Pozadzides 2010)

reported that the top referrers on the web as a whole were search engines (mainly Google), media sites (e.g., YouTube and Flickr) and social web sites (especially Facebook).

Web referral traffic is extremely important in the library domain, although except for Google, the top referrers differ. Students are aware of and have continued to rely on online scholarly sources, but they are now discovering them more often through Google, Google Scholar and Google Books (Hampton-Reeves et al. 2009, 36). Now that the content of scholarly aggregations (like ScienceDirect and the content of open access repositories) is crawled and centrally indexed by Google, a huge amount of traffic to online scholarly content comes from Google (CIBER 2009, 21; Hanson and Hessel 2009).

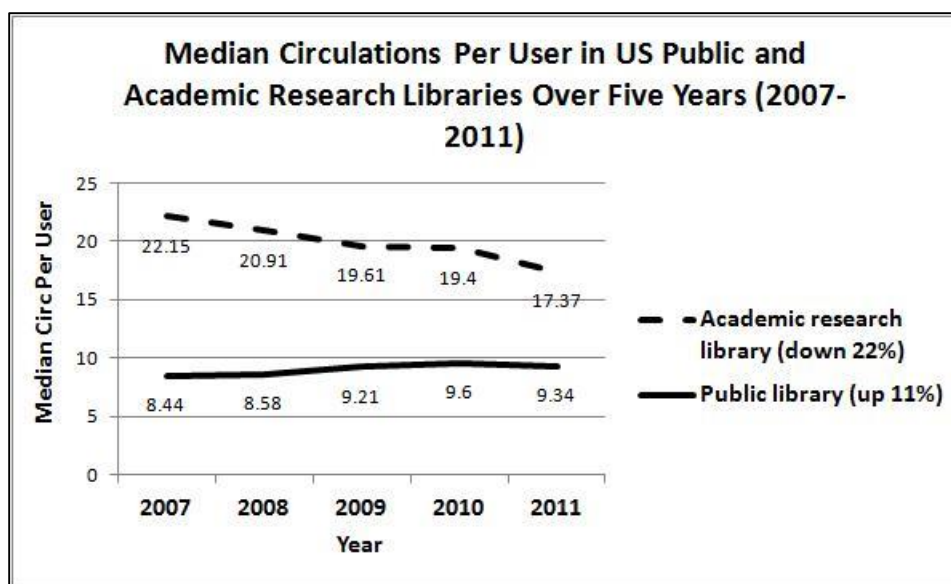
The US and UK Ithaka studies of 2012 suggest that for scholars, the most important role of the library is as a buyer/licensor of online content (US survey, 67-68; UK survey 79-80). This is not to say that libraries' provision of online catalogs and library web sites is no longer important—it is—but it is important to understand the context in which library catalogs and websites function in the larger web environment. Hanson and Hessel (2009, 26-28), in their groundbreaking “discoverability phase 1” report for the University of Minnesota Libraries, reported that 75% of the traffic to the libraries' reference linking service (enabling connections to library e-resources) originated from external referrers, specifically Google, PubMed and the web sites of scholarly databases or indexes.

### ***Changing use and engagement with hybrid libraries***

Since about the 1990s, the position and comparative use of traditional library collections have changed dramatically. Hybrid library users are increasingly finding and engaging with library materials on the larger web, rather than visiting library sites as often as before. This section uses data for US public and academic libraries to illustrate these trends.

### Comparative demand

The patterns of hybrid library collection use are different in academic and public libraries. There is a consistent downward trend from 2007 to 2011 in the circulation of the printed books and journals in ARL library collections ([arlstatistics.org](http://arlstatistics.org)). Data from the US Public Library Data Services (Reid 2012) indicates that circulation of public library collections (which contain high-demand popular materials) has shown an upward trend between 2007 and 2011 (figure 5-1).



**Figure 5.1 Trends in the Use of Public and Academic Research Library Traditional Collections**

Sources: Public Library Data Service (Reid 2012) and Association of Research Libraries ([arlstatistics.org/analytics](http://arlstatistics.org/analytics))

### Academic libraries

Academics demonstrate what they want by what they use. The academic library circulation trends for the physical collections are directly related to the findings of the user studies cited previously in this chapter. Academic research, teaching and learning increasingly relies on scholarly digital content and less on print.

Academic library print circulation trends are inversely correlated with the high traffic to the scholarly digital libraries like those in table 2-1 (e.g., the ACM Digital Library, JSTOR, ScienceDirect). Tripathi and Jeevan (2013) offer an extensive literature review of the many aspects of the usage of e-resources in academic libraries: usage statistics, analytical methods, usage patterns across disciplines and institutions, information-seeking strategies and the growing importance of assessment.

### *Public libraries*

US public libraries offer access to growing numbers of e-serials and scholarly databases, and state agencies typically purchase e-content licenses for the libraries in their states.

Notwithstanding the provision of access to e-serials and databases, public library user demand is centered primarily on books, audiovisual materials and increasingly, e-books. Public libraries in the US collect materials largely for popular use and for children—circulation of children's materials accounts for a third of total circulation, according to the IMLS Public Library Surveys of 2009 and 2010. Books (which account for about 85% of public library collections) also account for most of what circulates (63%). However the 2009 and 2010 IMLS surveys reported that audiovisual materials and e-books are the fastest growing components of public library collections, and e-book demand is growing dramatically (Reid 2012; Miller et al. 2011 and Swan et al. 2013; Bowles and Hazzan 2013; Hoffert 2013; Price 2013).

### *Demand for e-books*

After a long foreground that featured debates about issues with reading online, whether publishers should or should not offer e-books, and other issues, the public's use of e-books and the ownership of e-book readers and tablets are finally taking off. The timelines in figure 4-1 mark the points at which Google released a million public domain titles from the Google Books project in the EPUB e-book format (2009), Amazon e-book sales topped hardcover book sales

(2010) and ownership of tablets and e-book readers reached 20% of US adults (2012). Rising e-book demand in US public libraries has already been mentioned. In US academic libraries, e-books are taking off more slowly. Restrictive licensing terms, resulting in trouble downloading files or printing more than a few pages, and other problems continue to slow acceptance and adoption rates (Walters 2013a, 2013b).

#### *Demand for digitized special collections*

Chapter 2 and table 2.1 discuss some of the early digital libraries of cultural heritage content that attracted considerable use by scholars, teachers and citizens. National library digitization programs in particular have attracted attention and high use by new types of audiences. These programs digitized books but also images, sound recordings, newspapers and more.

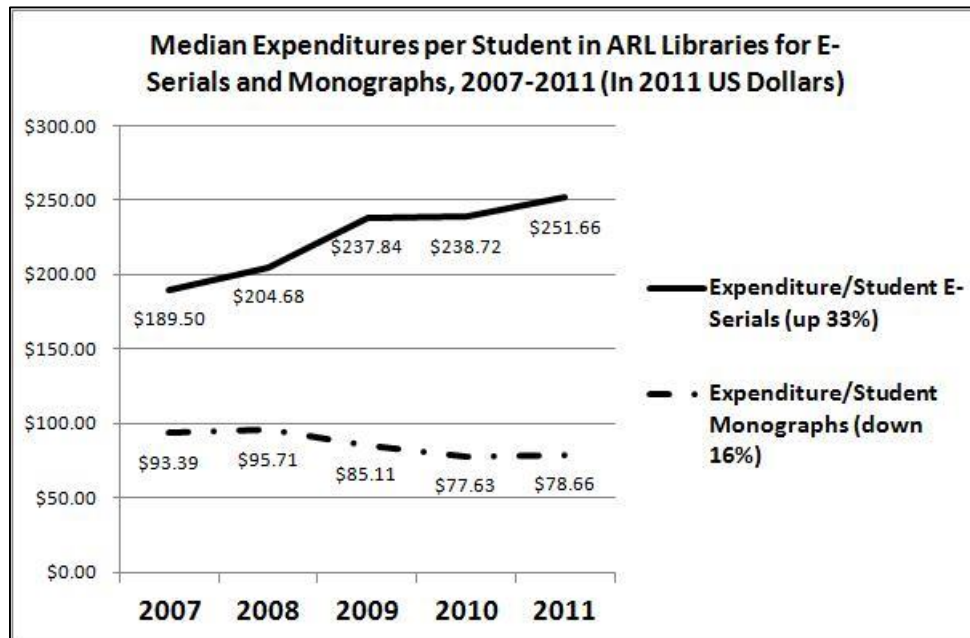
### **Libraries' response: changing hybrid library collections**

#### ***Expenditures of materials budgets***

Libraries tend to buy what their communities use. Hoffert (2013) explores changes in how US public libraries are spending their materials budgets based on the 2012 Library Materials Survey conducted by *Library Journal*. Survey results suggest that materials budgets are holding steady, and public libraries are spending 59% of their budgets on printed books, down from 68% in 2006. The difference appears to be going to media and e-books.

A combination of factors including rising demand for e-content and falling demand for print, combined with the economics of e-resource licensing (i.e., rising prices), has led to dramatic changes in how academic libraries spend their materials budgets. Based on median amounts, ARL libraries spent 42% of their materials budgets on e-serials in 2007, rising to 58% in 2011. For monographs, they spent a median of 21% of their materials budgets in 2007, dropping to

18% in 2011 (arlstatistics.org). Figure 5-2 provides another view of these expenditure trends over five years based on expenditures per student.



**Figure 5-2 Rising E-Serials Expenditures, Dropping Monograph Expenditures**

Source: Association of Research Libraries (arlstatistics.org/analytics)

### ***Managing physical collections in academic libraries***

In the face of changing usage patterns for printed books and serials, academic library leaders began to ask serious questions about their low-use print collections, especially in light of the space that such collections take up in library buildings. Many libraries were crowded with people wanting different sorts of services (e.g., group study space, computers/information commons, space for library instruction, more seating in general). These questions took on new intensity after about 2009, when a large corpus of mass-digitized books had emerged.

#### ***Print collection management***

David Block's article on the history of library collection storage notes that many US research libraries had reached their capacity for storing their collections by the 1980s (Block 2000). High



density storage was the first type of solution sought (like the Harvard Depository, which dates from 1986). By 2005, there were 50 or more library storage facilities in the US and others in the planning stage. Most housed individual library collections but shared storage facilities were beginning to appear (Payne 2005). Vattulainen (2004) reviewed the role of national or regional print repositories in Finland and several other European countries. In 2008 O'Connor and Jilovsky concluded their review of library collection storage solutions in a number of countries (UK, Australia, Finland, US, etc.) with a recommendation for a network of shared national or international print repositories. By about 2007 research library storage solutions had become a regular component of research library collection management strategy, and a component of preservation strategy as well (Rosenthal 2010).

#### *Conceptual, political or operational barriers*

As the physicality of library collections has become less important, and digital content becomes more plentiful, rich and diverse, a trend of rethinking library collection strategies began. At the same time, research libraries have continued to be reluctant to undertake storage of large parts of their collections for conceptual, political or operational reasons.

#### *Collections of e-books?*

It remains an open question whether e-book licensing will substantially replace print book collecting going forward. A major shift to e-books for providing access to current titles for academic libraries could happen, but many serious barriers remain (Walters 2013a, 2013b). So far the tidal wave-like adoption of digital formats for databases, journals and articles that occurred in academic libraries between 2000 and 2005 has not been repeated with e-books. Meanwhile, a number of research libraries are experimenting with an innovative method for licensing e-books, called Patron-Driven Acquisitions, PDA, a model for licensing and purchasing e-books “just in time,” based on library patron selections, rather than having librarians select

and buy them “just in case” they are needed (Nixon, Freeman and Ward 2010; Hazen 2011, 200; Fischer et al. 2012).

### *Rising priority of special collections and archives*

As more scholarly content moves online and academic libraries license the same or similar e-content packages, individual libraries’ online collections have become less distinctive. There is also considerable overlap in many legacy print collections (see section on mass digitization). Special collections and archives are what remain most distinctive about research library collections, and the results of cultural heritage digitization projects suggest that if such special collections were more discoverable online, they would attract new users and uses.

A number of reports recommend raising the priority of library efforts to enable the online discovery and use of special collections and archives (Loughborough University. Library & Information Statistics Unit and Research Information Network 2007; Association of Research Libraries 2008; Dooley and Luce 2010). Some research libraries have been able to digitize parts of these collections and produce new finding aids that make these archives more visible on the web, either through institutionally-funded projects or through partnerships (see for example Hawkins and Gildart 2010 and Bingham 2010 on the partnership to digitize British historic newspapers 1600-1900). As of this writing, however, many important collections of primary sources continue to be hidden and inaccessible to discovery on the web.

### ***Digitizing research library collections***

Chapter 2 and table 2.1 provide a number of examples of successful digital libraries that have their roots in the first decade of digital library work. These projects produced digitized library collections at small and large scales. One outcome of these early projects was to demonstrate the exciting potential, feasibility and value of digitization projects and techniques. The projects

were characterized by careful selection of materials to be digitized and the development and use of digitization best practices. Preservation was an element of many if not most projects. In late 2004, Google introduced a new approach called “mass digitization.” Mass digitization generally alludes to the digitization of very large, whole collections of content, with no or minimal selection.

### *The Google 5*

In December 2004, Google announced agreements with five major research libraries (the New York Public Library and the libraries of Harvard, Michigan, Oxford, and Stanford) that enabled Google to digitize volumes from these libraries’ printed book collections. These libraries were called the Google 5, and the project, which became known as the Google Books Library Project, now has more research library participants. The project has focused on indexing and access to book content and has no preservation component. The scale of the project and the speed with which it has progressed are unlike anything that came before it.

Lavoie, Connaway and Dempsey (2005) evaluated the Google 5’s collections to estimate the proportion of the system-wide book collection that they represent. Their results suggested that the combined Google 5 collections could potentially be 10.5 million books, with the following characteristics:

- They would represent 33% of the system-wide book collection at that time
- Half of the books would be in English, with another quarter of the remaining books in German, French and Spanish
- 80% of the books would be in copyright

*Other US-based projects announced in 2005*

In 2005 a second mass digitization project was announced, called the Open Content Alliance, with the goal of digitizing public domain books. The project had funding from Microsoft, Yahoo and others, and the scanning was done by the Internet Archive (Coyle 2006; Hahn 2008). That same year, Microsoft announced a mass digitization project called Live Search Books, another cooperative project with libraries; it ran from 2006 to 2008. The project's 750,000 digitized books are now part of the Internet Archive (Jones 2010). Also that year the Librarian of Congress announced that Google had provided US\$3 million to jumpstart the World Digital Library (see also [wdl.org/en/contributors](http://wdl.org/en/contributors); Hahn 2008).

*Quand Google défie l'Europe*

Jean-Noël Jeanneney, then President of the Bibliothèque nationale de France, responded about a month after Google's announcement with an editorial in *Le Monde*. It was called "Quand Google défie l'Europe" ("When Google Challenges Europe"; Jeanneney 2005). The editorial was later expanded into a book, *Google and the Myth of Universal Knowledge: A View from Europe* (Jeanneney 2006).

In his 2006 analysis of the situation, David Bearman, a digital library leader and the founder of Archives and Museum Informatics, took the position that Jeanneney "succeeded to a significant extent in motivating a movement to digitize European print heritage." Bearman's article provides an overview of Jeanenney's compelling critique of Google and the Google Library Project.

*i2010 Digital Libraries*

In September 2005 the European Commission announced i2010 Digital Libraries, a plan to build a European digital library containing six million digitized books and other materials by 2010

(Forster 2007). The initiative was intended to build on the organizational framework of TEL (The European Library), an initiative of CENL (Conference of European National Librarians).

The aims of the i2010 Digital Library initiative were lofty: to provide for the digitization, online accessibility and preservation of Europe's cultural memory. The approach has been to work with publishers and libraries on the intellectual property rights aspects of the initiative (including how to manage orphan works). The European initiative also has preservation objectives.

### *Europeana*

Europeana ([europeana.eu](http://europeana.eu)) is the digital library that grew out of i2010 Digital Libraries. A preliminary version of Europeana went live in late 2008, followed by the first operational version in summer of 2010. That version provided access to over ten million digital objects from libraries, museums, archives and audiovisual archives from across Europe (Chambers and Schallier 2010). At the time of this writing, Europeana provides access to 20 million digital objects. Chapter 10 discusses Europeana in more detail.

### *HathiTrust*

Large-scale digitization has the potential to transform the library world. The launch of HathiTrust Digital Library ([hathitrust.org](http://hathitrust.org)) in October 2008 created new momentum for such transformation. It began as a partnership of the Committee on Institutional Cooperation (CIC; a consortium of 15 mostly midwestern US universities) and the University of California System. Many new partners have since joined HathiTrust, which uses a membership model to fund its operations and services. It is not a commercial or government funded operation. HathiTrust's goals include creating a shared repository of digital collections for access and preservation and stimulating efforts for shared collection management strategies. The commitment to preservation is particularly strong.

Most of the HathiTrust repository consists of digitized content from libraries that participated in the Google Library Project. Other sources are digitized content from the former Microsoft Live Search Books project, the Internet Archive, and books digitized by the partners themselves (Christensen 2011). The HathiTrust has many services, among them mechanisms for reviewing and documenting copyright, APIs, and metadata that libraries can load into their online catalogs. In June 2013, the Digital Public Library of America (DPLA) announced a partnership in which HathiTrust will share its public domain content, representing some 3.5 million volumes, with DPLA.

#### *The lawsuit against HathiTrust*

In September 2011 the Authors Guild and others brought a suit against HathiTrust alleging that HathiTrust's storage and search of full-text digital books is an infringement of copyright. A court within a particular circuit of the US federal system heard the case. The court provided a ruling in November 2012 stating that HathiTrust's retention and use of digitized books for purposes of preservation, text search, and accessibility for the visually impaired are within the limits of the US laws regarding fair use (Crews 2012). Since HathiTrust has not acted on its preliminary plans to make orphans works accessible, the judge did not comment on whether HathiTrust's plan would have been lawful. Chapter 3 briefly discusses the legal issues associated with Google Books and the HathiTrust Digital Library.

#### *Implications for future collection management*

At the time of this writing, the HathiTrust digital collections contained close to 11 million digitized volumes. Europeana provides access to 20 million digital objects (including books). Despite the open legal issues around mass digitization projects, when Europeana, HathiTrust and other initiatives are considered together, it seems clear that the time has arrived when the content of

academic library books is no longer limited to those with access to the physical collections held in academic library buildings. This content is now online and abundant.

In 2011 Constance Malpas reported on a Mellon-funded project to study managing print collections in a mass-digitized environment. With participation from OCLC Research, HathiTrust, the library of New York University and the Research Collections Access & Preservation (ReCAP) consortium, the project investigated the feasibility of radically different solutions for managing low-use print books using large-scale, shared print and digital repositories .

At current digitization rates, the HathiTrust Digital Library is expected to duplicate 60% of the retrospective collections of ARL libraries by June 2014 (Malpas 2011, 10-11). If shared print and digital repositories were implemented, these research libraries could achieve significant efficiencies and repurpose thousands of square feet in their libraries for learning or information commons, media labs and other uses.

#### *Shared print repositories and mass-digitized books*

Robert Kieft and Lizanne Payne (2012) wrote an article that is cause for cautious optimism that new shared solutions are both practical and likely to emerge. They explore the concept of large-scale, regional and national cooperation for hybrid library collection management. They take as a given that the current legal obstacles around mass-digitized books will eventually be resolved through new legal models negotiated between libraries and publishers and new business models for compensating rights holders (p. 140). The first part of the article lays out a detailed vision for the collective management of hybrid library collections in the 2020s. The second part provides examples of new models, and the article closes with a suggested research agenda for “collective collecting.”

## **Changing technologies for hybrid libraries**

### ***Library management systems and business processes***

By the late 1990s, the current generation of library management systems (also known as integrated library systems) were being implemented. These systems consist of integrated software applications generally based on relational databases. Library management systems support the business processes (activities that produce services or products) of libraries: selecting, acquiring, describing and managing, discovering, circulating/delivering/linking, and preserving library collections plus evaluation. They generally have two interfaces: one for staff use and one for end users (the library online catalog).

These library management systems were initially developed at a time when library collections were still dominated by print. They have proved challenging to adapt to a world dominated by e-resources and new requirements for hybrid library collection building and management. This mismatch kicked off a technology replacement cycle that is still underway. At the time of this writing, my knowledge and evaluation of the landscape suggests that hybrid libraries are in a transitional period featuring many types of interim solutions.

### ***Technology replacement: in transition***

The enabling technologies of large academic libraries today are a complex, decentralized patchwork that stitches together various components. These components support hybrid library business processes on the one hand, and end-user discovery and access on the other. Achieving interoperability across this complex patchwork of enabling technologies is labor-intensive and costly, and for a variety of reasons some types of digital content (e.g., institutional

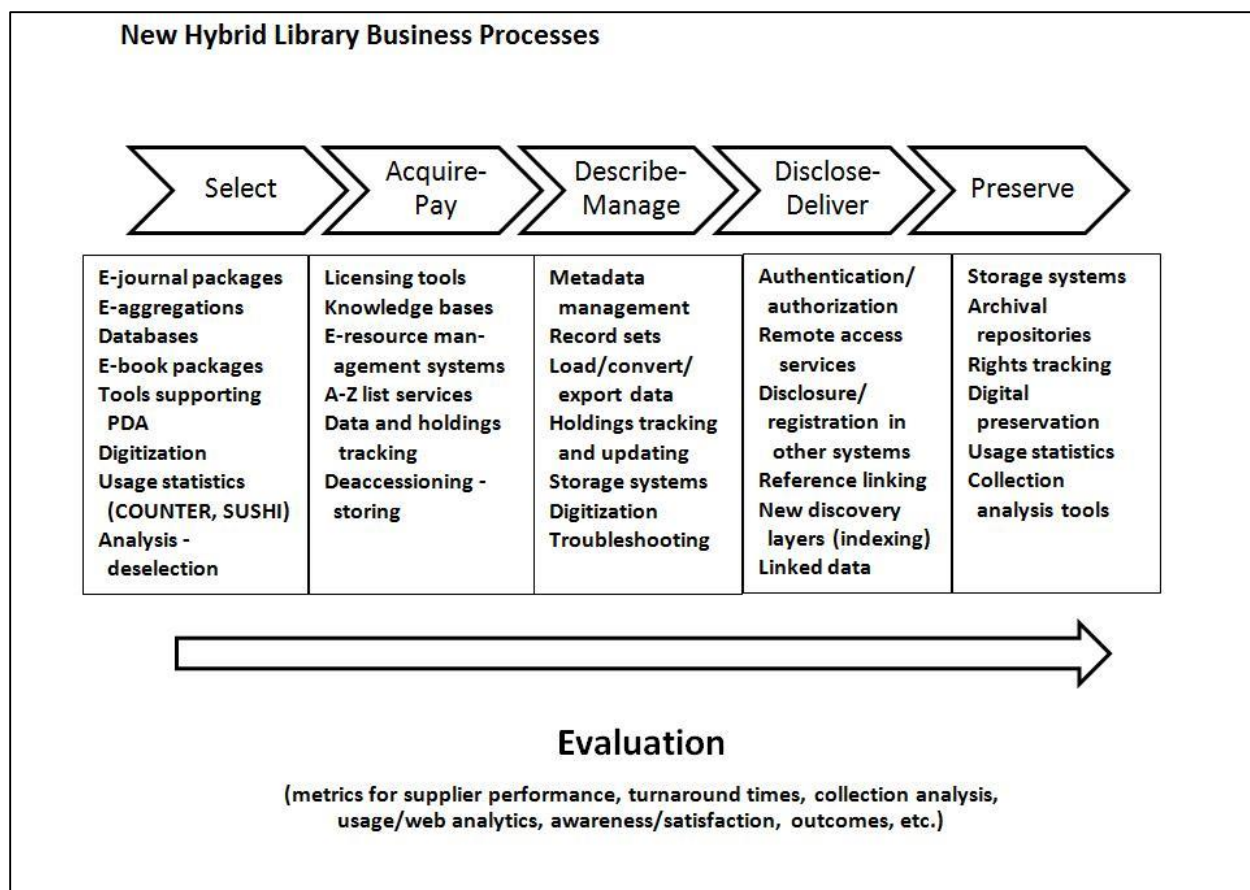


repositories, local digital library content) is often not integrated in the mix at all (Menzie, Birrell and Dunsire 2010).

Kress and Wisner (2013) offer an interesting model (based on supply chain management) for beginning to rethink and improve upon the current situation, but so far an overarching strategic framework for hybrid library enabling technologies—let alone an actual integrated solution—does not exist. Given the constant turbulence of the web environment, a new technical solution for managing hybrid library collections may not look much like library management systems up to now. Regardless, interoperability is a key challenge in hybrid libraries now and it will continue to be challenge going forward.

### ***Types of enabling technologies and tools***

Figure 5.3 illustrates the types of enabling technologies and tools now associated with the business processes of hybrid libraries. The business processes are listed along the top (select, acquire and pay, describe and manage, disclose and deliver, preserve), technologies and tools are in the text boxes below, and examples of evaluation processes are listed at the bottom of the figure. The figure labels the business processes as “new” not because the processes themselves are new but because the technical requirements and tools for supporting them are new.



**Figure 5.3 Technologies and Tools Supporting Hybrid Library Business Processes**

The reports of the second two phases of the University of Minnesota Libraries’ “discoverability” studies offer an interesting parallel to figure 5.3’s visualization of transitional hybrid library technologies and tools. Hanson and colleagues (2012) articulate a vision for a new discovery environment that (1) integrates content and metadata from different sources, (2) exposes content and metadata to external systems and services, (3) indexes content from external sources (e.g., HathiTrust), (4) allows for personalization, and (5) provides evaluative information to support user-centered, evidence-based decision making.

Fransen and colleagues (2012) report on the third phase of the discoverability studies includes helpful thoughts about system requirements, a drawing of their technology “ecosystem” as of

2011, and some comments on the cloud-based library management systems that are currently available.

It would take a whole book to describe all the technologies and tools in figure 5-3 and how they help to achieve the purposes of hybrid libraries. In recent years, hybrid libraries have invested a great deal of attention to how library management systems support the discovery of collections. The following sections describe this work, then turn to other efforts related to the progress of digital libraries.

### ***Interoperability and the problem of discoverability***

A key challenge for hybrid libraries is the same as the key challenge of digital libraries, discussed in chapter 3: interoperability (see chapter 3 for the full discussion of interoperability). An important objective of interoperability is discoverability, which involves integrating diverse digital content in a single system as well as making content discoverable in external systems and services. In the hybrid library context, discoverability has two dimensions:

1. **Disclosure and visibility of hybrid library collections on the network**, particularly on high-traffic sites. Study after study reported in this book and elsewhere provide strong evidence that for all types of people, information-seeking and discovery begins on web sites external to libraries. Individual hybrid library catalog data is generally not disclosed for crawling by search engines, and given the current redundant state of library catalogs, it would not make sense for search engines to crawl and index them. A better, network-level solution is needed for making library content discoverable in external systems and services, especially search engines. A later section returns to this topic.

2. **Institutionally- or consortially-based discovery services.** This type of discoverability has to do with integrating diverse content in a single site. Libraries have accomplished a great deal of progress on this dimension of discoverability in recent years, as discussed in the following section.

### ***Discovery services***

#### *E-resource discovery*

Because library management systems of the late 1990s were ill-equipped to do so, librarians began to work on supplemental methods to enable discovery and delivery of e-resources (databases and indexes, numeric files, full-text, etc.). Their first attempts using static web pages containing links and locally-created descriptions, then searchable databases quickly ran into problems of scale. Some early solutions featured “A to Z lists” providing links to the titles of e-resources from the library’s web site; these are now common services offered by vendors.

The provision of metadata sets for loading e-resource descriptions into library online catalogs is also a commonly offered service today. These sets exemplify the shift from title-by-title bibliographic control to automated metadata management in hybrid libraries. Increasingly, this automated approach is used to support selection, acquisition and cataloging of many types of content, including e-books. This kind of approach is also used to disclose or register and maintain library holdings to external systems so that hybrid library collections can be more visible on the larger web.

#### *Fragmented hybrid library interfaces*

Another strong motivation to seek unifying discovery methods on hybrid libraries’ destination sites was the proliferation of hybrid library interfaces: the library catalog, A to Z lists, static web pages, gateways, and more. Library users were obliged to know about these and search each

interface separately (Calhoun 2002, 149). Some of the separate interfaces continue to be needed, but libraries lacked one common user interface to everything—a single point of entry to their hybrid collections.

### *The promise of portals*

The term “portal” has a range of definitions, but from a functional perspective libraries wanted to simplify searching across and linking from and to diverse collections, and also make it easier to authenticate and authorize access to licensed resources.

### Authentication and authorization

When libraries licensed only a few databases and e-resource packages, it was possible to keep track of individual logons and passwords for each of the interfaces. Once there were hundreds of these, an automated solution to authenticating users and authorizing access to all of the resources became essential. Authentication is the automated process of identifying a person (often based on a user name and password), and authorization is the automated process of providing the appropriate access rights.

### Branding and a unifying interface

Portals were also intended to improve the library’s ability to “brand” itself as the provider of access to hybrid library content. Libraries wanted systems with a unifying interface that would federate searching of the distributed, heterogeneous content they licensed (e-content), wished to point to (publicly available web sites) or owned locally (non-digital collections); and they wanted the system to present the results in a coherent way to searchers. They also wanted to offer their communities the ability to link from an information object in one resource (for example a citation database) to an object in another (for example the full-article described in the citation).

Librarians referred to these functionalities as “metasearch” (also known as “federated searching”) and *reference linking* respectively.

### *The European Library (TEL)*

At the beginning of the new millennium many new projects got under way in libraries to explore the possibilities of portals. In Europe, early work went back as far as 1995, when the British Library and the national libraries of Finland and the Netherlands launched the pilot project GABRIEL (Gateway and Bridge to Europe’s National Libraries; Hakala 1999). That pilot provided experience that eventually led to The European Library ([theeuropeanlibrary.org](http://theeuropeanlibrary.org); TEL; Woldering 2004; Van Veen and Oldroyd 2004). TEL launched a new portal in 2005 and continues as a portal to collections as well as providing the channel for submissions of digital content to Europeana (discussed in chapter 10).

### *Problems with metasearch and Z39.50*

Other portal projects that tested metasearch were learning experiences that did not produce long-lasting services (see Feeney and Newby 2005; see also the annotated bibliography of Freund, Nemmers and Ochoa 2007 for further information about the problems of metasearch). By 2008, many early adopters of metasearch had replaced their implementations with other solutions (see Breeding 2012a).

### *A new kind of library catalog: discovery services and centralized indexing*

By 2005, it was clear that the traditional library online catalog was not going to be an adequate future discovery service (see for example Calhoun 2006, 38). There were too many new requirements that the current generation of library management systems, online catalogs and supplemental tools could not meet. The centralized indexing approach used by popular search

engines opened the market for new types of institutionally- or consortially-based discovery services.

The phrase “discovery service” has meaning in several contexts, for example among web developers. This book defines the phrase in a library context, where discovery services refer to user interfaces that provide for unified, integrated search and retrieval based on a pre-harvested, centralized index to heterogeneous resources. Typically the discovery service indexes the library’s licensed resources (e-journals, articles, e-books) and physical collections. Sometimes the index also points to external digital libraries (like HathiTrust). The service hosts the indexes centrally, and searchers get instantaneous results for their queries as the service links to and displays online full text. Discovery services are designed to meet the discovery and delivery requirements of hybrid library business processes (see figure 5-3). They do not address requirements for other business processes. Discovery services co-exist with library management systems. .

#### *Early discovery services*

Some institutions built discovery services early in the new millennium. A team at the libraries of Lund University in Sweden developed a discovery layer for e-content and launched it in late 2003 (Jørgensen et al. 2003; Mayfield et al. 2008). BASE (Bielefeld Academic Search Engine; base-search.net; Lossau 2004) anticipated the development of library vendors’ discovery services by five years or more and it is still thriving. As of this writing, BASE indexes over 48 million documents from more than 2,600 sources. In 2004, North Carolina State University’s librarians purchased Endeca’s Information Access Platform to create a new discovery layer and faceted search features for their library catalog (Antelman, Lynema, and Pace 2006).

AquaBrowser is an early service that offered a discovery layer based on visualization techniques and associative indexing. AquaBrowser was first launched in production in many public libraries in the Netherlands, and at the end of 2011 it had 250 installations (Breeding 2012a). These early implementations significantly advanced the field's thinking about revitalizing the catalog through the introduction of discovery services (see for example Lindahl, Bowen, and Foster 2007; Sadler 2009; Emanuel 2011). By late 2007, library service sector firms had introduced a number of discovery services (Sadeh 2007; Wilson 2007; Mayfield et al. 2008; The Library Corporation 2008).

#### *Evaluations of discovery services*

The amount of content indexed in a discovery service may be the most important feature for libraries; they want to be sure that those who use their discovery services can get to all the content they have so expensively licensed on behalf of the communities they serve. The functionality of the discovery interface is another key consideration. The library literature is now full of reviews and evaluations from librarians who have implemented one of these services. Some are Asher, Duke, and Wilson (2012), Fagan et al. (2012), Gross and Sheridan (2011); Holman et al. (2012); Stevenson et al. (2009); Stone (2010); Way (2010); Yang and Wagner (2010).

#### ***Next generation hybrid library systems (the cloud)***

Library management systems became less able to support the business processes of hybrid libraries as digital content moved to center stage. Cloud-based hybrid library systems may offer a better-integrated alternative to the current fragmented array of systems, tools and services that hybrid libraries must use. Breeding (2011, 2012a, 2012b, 2013a) provides highly readable information and annual updates on the emergence of cloud computing and cloud-based library systems.



By transitioning to cloud-based systems, libraries can replace their local library management systems with web-based applications that are accessed via common web browsers and whose infrastructure is supported “in the cloud.” There is no software to install or update, no local servers to purchase or maintain, and local maintenance activities (like nightly backups) are managed externally by the service provider. At the time of this writing, cloud-based systems are just beginning to be implemented for managing hybrid library collections.

### ***Licensing terms and conditions***

As e-resources and digital collections became major elements of hybrid libraries, it became necessary to know much more about the legal issues of licensing and providing access to them. Chapter 3 discusses the key challenge of intellectual property rights in digital libraries and the difficulties surrounding copyright. These issues manifest themselves in particular ways in hybrid libraries. The following sections provide a brief introduction to a couple of aspects of this large field of inquiry.

### ***Negotiating terms and conditions***

Libraries now license and purchase access to digital content (articles, e-journals, e-books) instead of purchasing the content itself. Publishers and other online information service providers restrict the rights to access, display and export most online scholarly content. Open access journals and repositories provide an alternative to licensed content, and they are helping to mitigate the asymmetrical relationship between publishers and licensees like libraries, but so far there is not a critical mass of open access scholarly content (see Hazen 2011, 198-200 for a discussion of this and other rights issues for research libraries). The problems are extremely complex and unlikely to yield to simple solutions. It is increasingly important for all librarians to have a basic grounding in the legal aspects of negotiating and adhering to the terms and

conditions of digital content licenses. Those who manage licensing in large research libraries need additional training and experience.

Knowing about licenses is important because much is at stake in terms of the library budget. North American academic research libraries now spend more than half their materials budgets on e-resources; in 2011, the median expenditure for this type of content in an ARL library was US\$7.3 million (arlstatistics.org). Most of the money (90% in 2011) is spent on costly bundles and packages of scholarly e-journals and articles. In 2012, ARL began tracking its members' expenditures on e-books; the median was US\$626 thousand per library.

#### *Licensing best practices*

Rachel Miller (2007) provides an excellent introduction to e-resource licensing best practices, education and training for licensing, model licenses and checklists, and key licensing issues (of which there are many). She briefly discusses licensing negotiation, the importance of tracking licenses and renewals, consortial licensing, pricing and cancellation terms, defining the population of authorized users, standard uses and fair use (e.g., for interlibrary lending and reserves), securing perpetual access rights to content, content loading and retention rights, copying for preservation purposes, and resisting non-disclosure agreements.

Best practices for licensing e-books are in an earlier stage of development, but librarians are carrying forward what they have learned about licensing e-journal packages. Lowry and Blixrud (2012) write that ARL libraries "did not want to repeat the license restrictions found in e-journal agreements that they are now trying to renegotiate." For example, Horava (2013) explores a variety of options and license models in the context of consortial licensing of e-books in Ontario, Canada.

### ***E-resource management, ERMs and e-resource usage metrics***

E-resource management has emerged as new specialization in hybrid libraries. The specialization matured quite quickly and now there are online discussion forums (e.g., LibLicense-L, [liblicense.crl.edu](http://liblicense.crl.edu)), workshops and educational resources, occasional and annual conferences (e.g., Electronic Resources and Libraries, [electroniclibrarian.com](http://electroniclibrarian.com)), and journals (e.g., *Electronic Library*). Enabling technologies and tools also emerged after about 2004 (Jewell et al. 2004; Ellingsen 2004; Fons and Jewell 2007), called “e-resource management systems” or ERMs.

E-resource management relies on knowledgebases, which are digital registries (machine- or human-readable, usually both) that collect and organize metadata and content needed for specific functions, like managing e-resource holdings, licensing and rights information. Another important enabling technology in the domain of e-resource management has been the collection of comparative e-resource usage data to support evidence-based decision making in libraries (COUNTER, SUSHI; see Chandler and Jewell 2006).

### ***Remote access to licensed e-resources***

#### ***Preference for remote access***

Enabling technologies were needed to manage who can have access and who cannot. The purpose of authentication and authorization mechanisms is to comply with the terms of licenses—but without requiring every user to log on for each session on each separate database or online full-text resource. For universities, often this was accomplished by giving the online content provider the institution’s range of IP addresses, which identify the computers or devices on its network. But this method of providing access did not work for authorized users who were connecting to the e-content from their homes and offices (this is called “remote

access”). Hanson and Hessel (2009, 25) found in their study of usage patterns at the University of Minnesota Libraries that 65% of requests for library online content came from off-campus. The marked preference for using e-resources from off-campus emerged early and is well documented in the US (Troll Covey 2003, 579).

#### *Enabling remote access to licensed e-resources*

The preference for remote access required another enabling technology to keep remote users' access from being blocked. In the US, hybrid libraries have provided for remote access largely with proxy servers or virtual private networks (VPNs). A proxy server intercepts remote users' requests and sends them to the server that delivers e-content. Remote users authenticate themselves by logging into the campus network. If authentication is successful the proxy server authorizes the remote user and passes along the request for content in a way that that “proxies” an acceptable IP address. Athens and Shibboleth are other popular authentication and authorization services used for managing remote access to e-resources. Even with these accommodations, troubleshooting e-resource remote access problems absorbs a great deal of the time and attention in libraries (Davis et al. 2012).

#### ***Disclosure and web visibility of hybrid library collections***

Progress in institutionally- or consortially-based discovery services is impressive. Progress on the other dimension of discoverability—disclosure and visibility of hybrid library collections at the network-level, on referring sites—is less noticeable. The evidence presented in this book suggests that a great deal of information-seeking for academic content has moved to search engines (especially Google), academic search engines like Google Scholar, or discipline-specific databases and aggregations. Discovery happens on these sites. The discovery to delivery loop is completed when the referring site sends the request to the appropriate server for delivering the e-content.

*Disclosure of e-resources in Google Scholar*

In his thoughtful article considering the impact of the introduction of Google Scholar in late 2004, Marshall Breeding proposed a serious reconsideration of the library community's approach to searching online resources (2005). Breeding predicted that Google Scholar might eventually become the default interface for finding scholarly information. The research reported in this book would suggest that it now has. For hybrid libraries, the success of Google Scholar implies that the disclosure and visibility of hybrid library collections in search engines and on other important referring sites is as important as the provision for institutionally- or consortially-based discovery of these collections.

*Representing libraries' physical and digital collections on the web*

Some cultural heritage digital libraries are reaching critical mass, gathering content from many contributors, so that they are popular destination sites on their own. For individual hybrid library sites, making their collections discoverable at the network level is crucial to their continuing value and relevance. Enabling technologies exist to allow disclosure and visibility of much scholarly e-content on top referring sites. More digital library managers are investing effort in improving the disclosure and visibility of repositories in academic search engines like Google Scholar.

The semantic web and linked data have promise for achieving greater disclosure of hybrid and digital library content, but at the time of this writing, few applications exist. Meanwhile, all the signs suggest that the technology associated with the discovery and reading of books is well into a new life cycle. Hybrid libraries need to make progress to heighten the discoverability of what they have to offer now. An encouraging development is the BIBFRAME project.

### *BIBFRAME*

The Bibliographic Framework Initiative ([loc.gov/bibframe](http://loc.gov/bibframe)) is a collaborative initiative of national libraries and other stakeholders, created and led by the Library of Congress since 2011, to examine the existing framework for bibliographic data and determine a transition path for MARC. The intent is to find ways to reap the benefits of newer technology like linked data. The project involves mapping the elements of MARC into a linked data structure. Zepheira, a firm with expertise in the standards of the semantic Web, the principles of linked data and web architecture, has been supporting the effort. In November 2012, the Library of Congress announced a draft data model for web-based bibliographic description, a primer introducing the data model (Miller et al. 2012) and a project to test its feasibility. The data model is called BIBFRAME. The *BIBFRAME Primer* (Miller et al. 2012, 28-38) describes other linked data initiatives as well as several other projects that have informed the development of the BIBFRAME data model.

### *Knowledgebases and registries*

An option is to move toward larger shared frameworks at the network level that would register many libraries' holdings and be able to feed this information to multiple providers on the web (Calhoun 2012b, under "Registries of library holdings"). The knowledgebases of cloud-based library management systems have the potential to provide the necessary metadata about holdings, etc. to such network-level registries, which would function to switch the user from discovery of content (on the network) to its delivery (from the appropriate library).

### *Some existing examples*

Approaches to interoperability based on registries and knowledgebases are already widely used on the web and on social sites. E-journal "A to Z" lists are produced from a knowledgebase that

registers and keeps track of the e-journals to which a particular library has access. ARROW is another kind of registry; its purpose is to facilitate rights management across the many libraries' digitization projects that are associated with Europeana (Caroli et al. 2012).

Google Scholar can already function, in practice, as a “registry” of library holdings for scholarly articles (given the relevant configuration data, Google Scholar switches user requests to connect to content to the appropriate institutional server). OCLC and Google Books have a partnership to provide a “Find in a library” service from Google Books. The Find in a Library service relies on library holdings data associated with the WorldCat bibliographic database. A study of the success rate of switching users from Google Books to a university library catalog found that the quality of the data supporting linking and delivery of content is as important, if not more important, as the data elements supporting discovery (Calhoun 2012a).

### *Engaging with the web*

Through consensus and partnerships, libraries could feasibly move to registry-based systems and make substantial advances in the discoverability of hybrid library collections on the web. These registries would automatically direct a user from sites where content is discovered to what his or her library holds (or licenses, or points to). A recent article by Lorcan Dempsey (2012) elaborates on these and other ideas for improving the visibility and utility of libraries on major network-level hubs. I have made the case elsewhere that the time has come for libraries to fully engage with the global network infrastructure, deploying methods that are native to the internet and web. Open access repositories have potential for becoming key building blocks in this process (Calhoun 2012b, under “Building for the web” and “Enlarged roles for open access repositories”).

## **Conclusion**

### ***Substantial and important progress***

Much of the output of scholarship is now online, and an increasing proportion of the books held in libraries are available in digital forms. Mass digitization projects have been incredibly productive, although legal barriers continue to block the full deployment of the digital content. The evidence presented in chapter 4 demonstrates that open access repositories are contributing more every year to the diffusion of scholarly content to the public. E-books are taking off in terms of supply and demand. National library projects and international programs like Europeana are making many cultural heritage materials open to all. In short, immense progress has been made toward an abundance of digital content.

A new specialization and enabling technologies for e-resource management have emerged and continue to develop. Librarians are getting better at understanding and negotiating licensing agreements with digital content providers. Some are also working proactively to advance favorable licensing provisions, broaden fair use for scholarship and for preservation, or expand open access. New discovery services have greatly advanced the discoverability of hybrid library collections on library sites.

## **Challenges**

### ***Enabling technologies***

While great progress has been made in enabling technologies supporting hybrid libraries, progress on the dimension of discoverability that relies on the adequate disclosure and visibility of hybrid library collections on the network is in its infancy. There is still no hybrid library management system that meets the requirements of hybrid library business processes (figure 5.3).



### *Awareness and relevance*

Figure 3-2 suggests that sustainability has several dimensions in addition to the economic one. The social dimension of sustainability involves visibility, community awareness and perceived relevance. Digital library and hybrid library collections must operate and demonstrate value in the context of today's web. Increasingly, information seekers do not need libraries to find scholarly digital content, digital books, and the many other types of content being created and consumed on the social web. Being a buyer or licensor of digital content is a tenuous claim to community relevance, subject to disruption by the next innovative business model for distributing this type of content.

### *The need for a collective strategy*

The parallel but separate evolutionary paths of repositories, hybrid libraries and the global web (see figure 4.1) are a major challenge for libraries going forward. Lavoie, Henry and Dempsey (2006) identified the risk posed by the lack of an overarching service framework for libraries some years ago and called for "reusable, recombinant, and interoperable library services." Reaching consensus about such a framework will require library cooperation and partnerships with a broad array of stakeholders.

A vision and examples of new shared solutions for managing hybrid library collections and mass-digitized books are beginning to emerge (e.g., Kieft and Payne 2012). These ideas for shared solutions for physical collections might be tested for their applicability to establishing coherent, collective strategies for repositories, hidden special collections, digitization programs, hybrid libraries and network-level interoperability.

*Demonstrating relevance and value*

Information seekers have many alternatives for getting the content they need. Relevance and value require more than being the owner of collections, a purchaser of licensed e-content, a publisher of open access journals, a creator or provider of digitized copies of cultural heritage materials, or the host of an open access repository or aggregation. Collections are not ends in themselves; they matter to the degree that they effectively support what libraries do for their communities, both local and global. The next chapter discusses digital libraries' social roles and their value to the communities they serve.